



L'interprétation en sémantique de corpus : le cas de la construction de terminologies

Anne Condamines

► To cite this version:

Anne Condamines. L'interprétation en sémantique de corpus : le cas de la construction de terminologies. *Revue Française de Linguistique Appliquée*, 2007, XII (1), pp.39-52. hal-00606198

HAL Id: hal-00606198

<https://hal.science/hal-00606198>

Submitted on 5 Jul 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

L'interprétation en sémantique de corpus : le cas de la construction de terminologies

Anne Condamines

CLLE-ERSS, CNRS et Université Toulouse Le Mirail

Résumé : Cet article tente de mettre en évidence la nécessité d'un double balisage pour mener à bien les analyses sémantiques à partir de corpus. Une première balise est constituée par la situation de production des textes et une autre par celle d'interprétation de ces textes. Pour les deux cas, l'auteur propose de mettre en œuvre la notion de genre (genre textuel et genre interprétatif) pour regrouper et catégoriser les situations. L'article s'appuie sur l'exemple de la construction de terminologies (relevant d'un genre interprétatif) en montrant comment le genre textuel influence le fonctionnement des marqueurs de relations conceptuelles (par exemple, la préposition *avec* pour repérer une relation méronymique). Il montre que ce type d'analyse permet souvent d'affiner des descriptions initialement faites par introspection.

Abstract : The aim of this paper is to focus on the necessity of a double marking-out when doing the semantic analysis of the data of a corpus. The first mark lies in the situation in which texts are produced and the second one lies in the interpretation of the texts. In both cases, the author suggests to use the notion of genre (textual genre and interpretative genre) in order to classify and categorize situations. The issue is exemplified by the problem of the building of terminologies according to a particular interpretative genre. The paper shows how textual genre influences the functioning of conceptual relations patterns (e.g., the preposition *avec* is used to spot a meronymic relation). It demonstrates that this kind of analysis may help to refine the descriptions initially made by introspection.

1. Introduction

La possibilité d'accéder à des textes en très grande quantité via l'internet ou les intranets (sites d'entreprises ou d'organismes) met le linguiste (et le sémanticien en particulier) dans une position paradoxale. D'une part, l'accès à ces données peut le décharger de l'obligation de constituer son observable à partir de sa propre compétence de locuteur (comme c'est le cas dans les approches introspectives). Mais d'autre part, la quantité de données accessibles produit un effet de masse qui

semble difficilement appréhendable. Dans un tel contexte, la dimension quantitative des phénomènes (nombres d'occurrences de telle ou telle forme) prend un sens et constitue un point de vue nouveau pour la description des phénomènes linguistiques. Mais des analyses fines, qui interrogent en profondeur le fonctionnement sémantique ne peuvent se satisfaire du seul comptage des formes et requièrent un renouvellement des points de vue. Il s'agit alors d'étudier la possibilité d'ordonner les phénomènes en mettant en place un balisage des études afin de dégager d'éventuelles régularités de fonctionnement. L'un des modes de balisage les plus travaillés fait intervenir la notion de « genre textuel » qui permet de dégager des catégories de textes en prenant en compte les influences réciproques de la situation de production des textes et des régularités langagières dans ces textes. Un autre balisage est beaucoup moins souvent mentionné mais il pourrait jouer un rôle tout aussi important dans la description de la variation dans les fonctionnements sémantiques, il s'agit de l'objectif de la description, c'est-à-dire de la situation d'interprétation et plus précisément de la possibilité d'identifier des catégories d'interprétation.

Le fait que l'analyse des textes soit elle-même située est depuis longtemps pris en considération, en particulier par les tenants de l'approche herméneutique (Rastier 1987 ; Salanskis & *al.* 1991). Mais la nécessité de tenir compte de l'objectif d'analyse a pris une importance accrue du fait même de la multiplication des textes disponibles. En effet, cette abondance s'accompagne de la conscience d'un besoin de mise en ordre dans les modes d'analyse de textes, tout particulièrement dans les organismes producteurs, qui sont confrontés à des problèmes cruciaux de gestion de leur documentation.

C'est la définition de ce double balisage de la sémantique de corpus (par la situation de production des textes et par la situation d'interprétation) qu'examine cet article en prenant l'exemple de la construction de terminologies. Bien que cet objectif ait une dimension applicative, l'article s'attache à montrer que, de manière plus générale, le renouvellement des demandes d'analyse oblige les linguistes à des descriptions qui, loin de n'avoir qu'une dimension appliquée (souvent seulement comprise comme définition de résultats ad hoc) permettent au contraire d'affiner les résultats obtenus par introspection et ouvrent des perspectives très prometteuses.

Le paragraphe suivant présente la sémantique introspective et la sémantique de corpus en insistant sur les éléments qui les caractérisent, soit conjointement, soit en opposition. Le paragraphe 3 situe la problématique qui sert de fil conducteur à cet article : la construction de terminologies à partir de textes. Le paragraphe 4 présente la façon dont le genre textuel intervient dans la description de deux marqueurs de relation conceptuelle.

2. Sémantique introspective vs sémantique de corpus

La possibilité d'accéder à des corpus met en question le rôle de l'introspection en sémantique. Alors que l'introspection joue un rôle déterminant dans les approches classiques de sémantique lexicale en mettant en jeu la mémoire lexico-sémantique du linguiste, elle pourrait sembler absente d'une analyse à partir de corpus dont on pourrait penser qu'elle n'a plus, quant à elle, qu'à organiser une réalité qui lui est

fournie sous la forme d'attestations. Pourtant, nous verrons que l'introspection n'est en aucun cas absente d'une analyse sémantique à partir de corpus mais qu'elle est déplacée sur d'autres phénomènes comme la nécessité de rapprocher des textes *a priori* grâce à la notion de genre.

Fondamentalement, tout linguiste, quel que soit son domaine d'intérêt, vise à identifier des régularités de fonctionnement qui soient reproductibles, enseignables voire prédictives. Nous partons en tout cas de ce postulat dans le présent article. Mais cette recherche de régularités se décline différemment selon les approches et de nombreuses caractéristiques opposent sémantique de corpus (dans laquelle nous englobons la sémantique textuelle) et sémantique introspective.

Objet d'étude

Pour la sémantique introspective, l'objet d'étude est majoritairement la phrase ou/et le mot, qu'il soit grammatical ou lexical. La sémantique lexicale est particulièrement représentative de cette approche qui s'interroge sur le sens des mots et les phénomènes de relations sémantiques entre les mots en langue, c'est-à-dire dans le système (synonymie, polysémie...).

La sémantique de corpus considère que son objet d'étude est le texte. Parfois, seulement le texte en tant que tel sans interrogation sur sa position par rapport aux autres textes (certaines formes d'analyse de discours, d'analyse littéraire, d'ethnolinguistique) ou parfois avec ce type d'interrogation.

Méthode

Le sémanticien introspectif convoque pour ses analyses sa propre compétence de locuteur, considérée comme assez bien représentative de la compétence du locuteur moyen. Il s'agit en fait de faire appel, même inconsciemment, à l'ensemble des attestations qu'il a rencontrées pour essayer de les catégoriser. Un des éléments pour lesquels cette intuition est régulièrement prise en défaut est celle de la répartition quantitative des phénomènes, l'intuition que l'on peut avoir de cette répartition correspondant rarement à la réalité des phénomènes en corpus.

La sémantique de corpus substitue à cette convocation mémorielle l'utilisation d'attestations réelles. La mise en ordre de ces attestations, loin de ne constituer qu'une manipulation neutre de données suppose au contraire de s'interroger sur leur mode d'interprétation. Le problème de la catégorisation se pose dès le choix des textes à constituer en corpus et nécessite un recours à l'introspection, au moins dans un premier temps.

Rôle du corpus

Le recours à des attestations est, pour le sémanticien par introspection, une façon de confirmer ses intuitions. Il s'agit par exemple d'utiliser des corpus électroniques comme *Le Monde* ou l'internet pour vérifier les descriptions réalisées par introspection (par exemple le type de construction de tel ou tel verbe). L'utilisation de corpus n'est ainsi pas totalement absente de l'approche introspective mais elle n'intervient que dans un second temps.

Pour la sémantique de corpus, la situation de production des textes étudiés est d'emblée problématisée et intégrée à l'analyse.

Ainsi, plusieurs caractéristiques sont communes aux approches introspectives et aux approches à partir de corpus : la nécessité de catégoriser (des mots dans un cas, des textes dans l'autre) et le rôle de l'introspection, c'est-à-dire du recours à notre expérience mémorielle de locuteur, justement pour effectuer des catégorisations. En revanche, deux éléments semblent irréductibles : l'objet d'étude (mot ou phrase vs texte), la prise en compte de la situation (avant et pendant l'étude dans la sémantique de corpus, après l'étude, pour vérification, dans la sémantique introspective) (Condamines 2005a).

Nous pensons toutefois, avec S. Bouquet (2004, 8), qu'une notion pourrait permettre de ne plus opposer aussi radicalement ces deux approches, celle de genre textuel :

[...] le genre, redéfini comme concept sémantique, peut faire lien entre les deux grandes traditions historiques de description du sens : la tradition rhétorico/herméneutique (qui prend en considération le contexte du sens, à savoir une détermination exercée par le palier de la globalité d'un énoncé sur le palier de ses composants locaux) et la tradition logico-grammaticale (qui ne rend compte que des « parties détachées de sens » décrites par des grammaires, formant systèmes et systèmes de systèmes, entrant dans la construction du sens selon un principe de compositionnalité [...]).

Pour resituer cette citation dans la perspective qui nous intéresse, on peut penser que le genre, parce qu'il fait intervenir une classification des situations de communication (c'est-à-dire aussi des textes) pourrait servir de palier de description qui autoriserait des classifications de mots (de sens de mots). Ces classifications pourraient être faites à partir d'attestations réelles. C'est bien sûr, donner un rôle très important au genre textuel dont on sait bien qu'il peut être lui-même très fluctuant. Mais il peut jouer un rôle pour comprendre l'articulation entre textes et mots, entre introspection et attestations.

2.1. Une sémantique doublement située

Afin de caractériser un balisage qui permettrait de dégager des régularités de fonctionnement, nous proposons de décliner ce balisage en prenant en compte deux moments clés : celui de la production des textes et celui de leur interprétation.

2.1.1. Des textes situés

Lorsque l'on souhaite baser une analyse sémantique sur la prise en compte d'attestations réelles, on est très rapidement amené à s'interroger sur le rôle du contexte de rédaction dans l'interprétation des faits. En effet, les textes ne sont pas seulement des attestations langagières mais des discours qui ont été produits dans des situations particulières.

La linguistique de corpus l'a démontré, un corpus n'est pas un ensemble d'attestations linguistiques mais un ensemble d'attestations que l'on a décidé d'examiner conjointement parce que ce rapprochement a un sens *a priori* (Habert & al. 1997). Dans cette élaboration, la prise en compte de la situation extralinguistique joue un rôle majeur : les textes qui sont rassemblés doivent avoir été produits dans des situations similaires. Cette affirmation appelle deux

remarques. D'une part, elle sous-entend que la situation extralinguistique est partie prenante de l'interprétation, c'est-à-dire de la construction du sens. D'autre part, la similarité des situations n'est pas un élément facile à caractériser : jusqu'à quel point peut-on neutraliser des variables de temps, de lieu, de niveau d'expertise, de domaine... ? En fait, bien souvent, la définition des textes pertinents (et donc de leur similarité) pour telle ou telle description linguistique constitue un résultat d'analyse, au même titre que la description des fonctionnements en elle-même. Pourtant, il faut partir d'une intuition sur le fonctionnement des textes, sur les éléments situationnels pertinents, pour arriver à cerner les phénomènes. C'est bien alors que réapparaît l'intuition, que l'on pouvait espérer voir disparaître dans une analyse à partir de corpus.

La prise en compte de cette imbrication entre extralinguistique et linguistique est depuis longtemps travaillée grâce à la notion de genre textuel. Longtemps cantonné à l'analyse littéraire (la notion existe depuis Aristote au moins (Genette, 1986)), le genre textuel est devenu, avec le développement de la linguistique de corpus, d'une urgente actualité (Swales, 1990).

Le genre textuel peut être vu sous trois points de vue : descriptif, prédictif, explicatif :

- descriptif, car il apparaît souvent qu'il y a une pertinence à décrire les régularités linguistiques en lien avec des situations extralinguistiques ;
- prédictif, car les descriptions sont censées être suffisamment fiables pour pouvoir décider qu'une situation extralinguistique « similaire » va produire les mêmes phénomènes langagiers ;
- explicatif, si l'on peut comprendre comment la situation et les phénomènes langagiers sont dépendants les uns des autres. Si cette compréhension est possible, il est évident que le pouvoir prédictif s'en trouve accru.

2.1.2. Une interprétation située

La prise en compte de la situation de production est revendiquée depuis longtemps par des disciplines comme la sociolinguistique, l'analyse de discours, l'ethnométhodologie... Mais il faut bien reconnaître que dans certains cas, la situation de production n'est pas suffisante pour expliquer des variations d'interprétation et qu'il faut aussi travailler la situation d'interprétation. Certes, les variations d'interprétation sont connues, par exemple en analyse littéraire, où elles sont parfois assimilées à des différences de réception individuelles. Mais nous voulons parler de variations qui ne sont pas liées à des individus mais bien à des locuteurs en tant qu'ils ont un point de vue interprétatif commun, souvent parce qu'ils sont engagés dans une tâche commune. Ainsi, les notices d'utilisation sont destinées à des locuteurs qui veulent savoir comment employer tel ou tel appareil et on peut supposer qu'ils vont décoder les textes explicatifs de la même façon. Mais parfois, les textes sont « détournés » de leur objectif premier et mis en œuvre pour des études particulières, à visée tout aussi collective : traduction, construction de définitions, archivage, recherche d'information... Ces situations sont, pour le linguiste, particulièrement intéressantes. En effet, elles constituent une sorte de laboratoire de ce qu'est l'analyse sémantique. Dans ces cas-là, l'objectif ne peut pas être (ou pas seulement être) la recherche d'un message que le locuteur est

censé avoir transmis ; il s'agit toujours d'une construction (dont on n'est jamais sûr qu'elle soit une re-construction). Ces situations d'interprétations particulières obligent ainsi à une analyse qui relève bien de la sémantique mais une analyse d'emblée appréhendée comme une construction et pas seulement un décodage. Nous proposons ainsi d'appeler *genre interprétatif* cette activité interprétative qui présente de fortes similarités avec le genre textuel. Différents arguments plaident en faveur de la définition de genres interprétatifs. Outre le fait que différentes disciplines comme l'ingénierie des connaissances ou les sciences de l'information tiennent compte de l'objectif de leurs analyses, des éléments plus théoriques peuvent argumenter en faveur de l'existence de genres interprétatifs.

Le principal argument théorique à cet égard repose sur le constat que la construction du sens est bien trop souvent considérée du seul point de vue du locuteur ou du rédacteur. Mais celui qui reçoit ou utilise la parole, interlocuteur reconnu comme tel ou autre type d'interprète, est également impliqué dans un processus de construction du sens. Ainsi, la situation d'interprétation, tout comme la situation de production, relève d'un processus d'élaboration sémantique. S'il en va ainsi, le même type de questionnement se pose pour la situation d'interprétation que pour la situation de production : comment le processus sémantique peut-il être à la fois individuel et collectif ? Tout comme on a proposé la notion de genre pour expliquer l'inscription collective d'une production textuelle, il semble intéressant et justifié de proposer cette même notion dans le cadre de la situation d'interprétation. Tout comme la situation de production donc, la situation d'interprétation pourrait relever d'un genre, c'est-à-dire d'un paradigme préexistant, voire « normatif » et en partie inconscient dans lequel s'inscrirait toute activité interprétative, ces trois éléments (inscription collective, caractère « normatif » et phénomène en partie inconscient) correspondant aux éléments principaux qui décrivent le genre textuel. Cette hypothèse signifie à la fois que, à cause de sa nature essentiellement sémantique, l'interprétation ne peut être strictement individuelle et que les points de vue interprétatifs peuvent être regroupés. Tout comme pour les genres textuels, il paraît difficile de dresser une liste figée de genres interprétatifs, mais tout comme pour les genres textuels aussi, certains points de vue interprétatifs semblent consensuellement acceptables, de manière intuitive.

Finalement, nous pensons que la sémantique de corpus peut être balisée par deux éléments : la situation de production et la situation d'interprétation. Ce double balisage, loin de ne constituer qu'une contrainte stérilisante, permet au contraire d'affiner des descriptions qui ont été élaborées sur des bases introspectives. Nous allons particulièrement nous intéresser à la construction de terminologies.

3. Construction de terminologies à partir de textes

La terminologie textuelle (ou construction de terminologies à partir de textes) est relativement récente (une quinzaine d'années tout au plus). En effet, pendant longtemps, la terminologie a été considérée comme une activité à visée normative, censée garantir une communication transparente entre locuteurs d'un même domaine. Cette vision, particulièrement développée par Wüster dans les années 30, dans la perspective de la vision positiviste du cercle de Vienne, a été longtemps

prédominante (Slodzian, 2000). Soutenue par un point de vue idéologique (la recherche d'une amélioration espérée de la communication) et financier (la constitution d'une terminologie de référence serait constituée une fois pour toutes), cette vision a pourtant été battue en brèche par la conjonction de différents éléments :

- du côté des utilisateurs, la méconnaissance parfois totale de l'existence de ces normes par les ingénieurs censés les utiliser et leur inadéquation avec la réalité des usages dans les entreprises n'a pas contribué à les diffuser ;
- le développement de la gestion électronique des documents (archivage, recherche et extraction d'information, aide à la traduction...) a montré les limites d'une terminologie figée une fois pour toutes ;
- l'avènement de la linguistique de corpus a ouvert des possibilités de prise en compte des usages réels.

Autant d'éléments conjoncturels qui ont obligé à une remise en cause radicale de la nature des terminologies qui ne peuvent plus être constituées indépendamment de la réalité des usages et de la réalité des besoins. Dans le même temps, la terminologie s'est rapprochée d'autres disciplines, qui partagent avec elle le mode de représentation de la connaissance (représentation sous la forme de réseaux de concepts) et le constat de la nécessité de tenir compte de la réalité des usages. C'est le cas par exemple des sciences de l'information (thesaurus), de l'ingénierie des connaissances (ontologies) ou du traitement automatique des langues (ressources lexicales spécialisées). Ce rapprochement a permis la mise en place d'une réflexion pluridisciplinaire sur les modes de constitution possibles de terminologies en fonction de besoins clairement identifiés (Aussenac-Gilles & Condamines 2004).

3.1. Marqueurs de relations conceptuelles

Concrètement, une ressource terminologique se présente sous la forme d'un réseau de termes. Lorsqu'on l'élabore à partir d'un corpus, il faut donc trouver dans ce corpus d'une part les termes et d'autre part les relations entre ces termes. Une des manières efficaces de retrouver des relations consiste à faire appel à des marqueurs de relation. En linguistique, on trouve les premières traces de cette notion il y a vingt ou trente ans chez des sémanticiens comme Lyons (1978, « formulae ») ou Cruse (1986, « diagnostic frames »). Les travaux sur ces marqueurs se sont multipliés, tant dans une perspective de description linguistique que dans une perspective de traitement automatique. En effet, la possibilité d'utiliser ces éléments langagiers, de toutes natures (typographique, lexicale, grammaticale...) intéresse particulièrement les informaticiens qui voient en eux la possibilité d'aider à la construction automatique de terminologies. Mais souvent, les études se concentrent sur une vision quantitative, qui tolère beaucoup de bruits. En effet, les contextes susceptibles de contenir une relation sont toujours soumis à l'utilisateur qui sélectionne ceux qui lui semblent pertinents. De ce fait, les travaux en TAL (Traitement Automatique des Langues) ont peu approfondi le rôle du contexte. Le contexte est ici à entendre à la fois du point de vue linguistique (co-texte) et du point de vue extralinguistique (situationnel). Des analyses plus fines sont proposées par les linguistes, en particulier sur la prise en compte du co-texte. Les

exemples que nous présentons s'intéressent à la prise en compte du contexte extralinguistique en interrogeant le rôle du genre textuel.

4. Prise en compte du genre textuel dans l'analyse des marqueurs de relation

Nous l'avons vu, le genre textuel peut avoir un rôle prédictif pour les fonctionnements linguistiques. C'est vrai pour certains marqueurs de relation pour lesquels la prise en compte du genre textuel ouvre des perspectives très intéressantes non seulement du point de vue du traitement automatique mais aussi du point de vue de la description linguistique. Ainsi, *chez* qui permet dans des textes des sciences naturelles, à visée didactique, de repérer des méronymes (Condamines, 2000) comme dans :

(1) *Chez les colobinés, le nez fait saillie sur la lèvre supérieure.*

(méronymie entre nez et colobinés)

C'est le cas aussi de *avec* et de la reprise par anaphore infidèle dont le fonctionnement est décrit ici.

4.1. Le cas de *avec*

Toutes les descriptions de *avec* mentionnent son rapport possible avec la méronymie (Cadiot 1997 ; Choi-Jonin 1995) :

(2) *Une robe avec des dentelles*

est ainsi une robe qui possède des dentelles (dont une partie est constituée par des dentelles).

Ces descriptions mentionnent aussi que la partie ainsi exprimée n'est pas une partie « essentielle » de l'objet : toutes les robes n'ont pas des dentelles et n'en sont pas moins des robes.

Une recherche menée sur des corpus permet de préciser le fonctionnement de ces *avec* méronymiques et, dans certains cas, de le remettre en question (Condamines, 2006).

D'un point de vue quantitatif tout d'abord, une étude menée sur des textes relevant de genres différents met en évidence que cette valeur méronymique est très inégalement répartie.

Cinq textes, relevant de genres différents ont été étudiés :

- un roman de Zola (Germinal),
- un manuel de géomorphologie, noté GEO
- un catalogue de jouets (catalogue de jouets Leclerc, Noël 2002),
- un ensemble de petites annonces immobilières (sélectionnés dans trois sites différents, de trois villes différentes), (P.A),
- un ensemble de descriptions d'itinéraires (12 itinéraires énoncés par 40 locuteurs : Itiné.).

Voici les résultats quantitatifs de cette étude :

	Germinal	GEO	Catalogue	P.A	Itiné.
Nbre de mots	209 200	219 100	9 200	12 630	48031
Nbre de <i>avec</i> (rapport avec le nbre de mots)	667 0,31%	432 0,19%	236 2,56%	185 1,46%	116 0,24%
Nbre de <i>avec</i> méronymiques (rapport avec le nbre de <i>avec</i>)	43 6,45%	55 12,73%	161 68,22%	141 76,22%	75 64,6%

Tableau 1. *Résultats quantitatifs dans l'étude de avec méronymique*

Il apparaît que la valeur méronymique est nettement dominante pour les trois derniers textes (catalogue, petites annonces, description d'itinéraires) par rapport aux deux premiers (roman et manuel). Ces résultats sont particulièrement parlants mais il convient d'essayer de les comprendre.

4.1.1. D'un point de vue qualitatif, en analysant de plus près ces *avec* méronymiques, on se rend compte de fonctionnements assez particuliers. En effet, lorsque les *avec* méronymiques sont nombreux dans un corpus, ils n'ont pas seulement une valeur informationnelle concernant une partie mais ils semblent plutôt jouer le rôle de soutien à une autre valeur, que nous avons qualifiée de saillance. Selon les corpus, cette saillance se décline différemment : saillance commerciale dans le cas des corpus à visée commerciale (catalogue et petites annonces)

(3) *Porteur évolutif avec canne, repose-pieds et ceinture de sécurité amovible.*

(4) *A l'étage : 3 chambres avec placards.*

et saillance perceptive dans le cas des descriptifs d'itinéraires :

(5) *C'est un pont avec des lampadaires en ferraille verte.*

Certes, c'est à chaque fois une partie qui est introduite mais c'est une partie qui donne un relief particulier à l'objet : une valeur qui justifie le prix dans un premier cas, une valeur qui, par son extraordinaire, permet de repérer un édifice dans le second cas. Cette hypothèse semble confirmée par deux éléments. Dans le cas des petites annonces immobilières, il arrive assez souvent qu'en alternance avec *avec*, on trouve le signe + (dont la valeur positivante est évidente) :

(6) *Villa [...] avec hall de nuit + placards et penderie.*

Dans le cas des descriptifs d'itinéraires, la valeur méronymique disparaît parfois au profit de la seule valeur de saillance perceptive comme dans :

(7) *Vous allez traverser un espèce de mini-parking, avec un plus grand parking sur la gauche.*

Enfin, certains exemples remettent en question la notion d'essentialité de la partie introduite par *avec*. Il s'agit d'exemples extraits du catalogue de jouets :

(8) *Cuisine avec plaque de cuisson qui rougit vraiment.*

Dans les extraits de ce type, il est clair que les objets sont décrits comme s'il s'agissait d'objets pour adultes et il est important que les parties décrites soient

conformes aux parties de l'objet stéréotypé. L'utilisation fréquente de *vrai*, *vraiment*, *réellement* a pour but d'augmenter cet effet de réel. Les parties introduites par *avec* sont ainsi les parties considérées comme essentielles pour l'objet.

4.1.2. Si l'hypothèse du rôle explicatif de cette notion de saillance est pertinente, alors, on peut aller au-delà de la seule description pour envisager une perspective prédictive : lorsque la situation correspondra à une saillance commerciale ou perceptive et que la méronymie y sera particulièrement présente, alors, on peut supposer que les méronymes seront massivement introduits par *avec*. A titre de vérification, nous avons recherché dans internet des sites touristiques dont on peut faire l'hypothèse qu'ils s'inscrivent dans une situation dans laquelle à la fois la saillance commerciale et la saillance perceptive (visuelle) sont pertinentes. Et en effet, dans les deux sites que nous avons consultés (l'un sur la Provence, l'autre sur le Pays basque), nous avons retrouvé une utilisation massive de ces *avec* méronymiques :

- 16/31 *avec* méronymiques dans le site de la Provence
(*le port avec sa grève*),
- 11/19 *avec* méronymiques dans le site du Pays Basque
(*Eglise avec un portail Renaissance*).

Toutefois, il faut prendre ces constatations avec beaucoup de prudence. En effet, une autre préposition apparaît pour la méronymie (dans des proportions bien moindre cependant), la préposition *à* (*tour carrée aux murs blanchis, vastes plages au sable fin*).

C'est aussi cette préposition qui apparaît de manière quasi exclusive dans les « titres » de plats : *omelette aux champignons, flan au potiron*.

On aurait pourtant pu attendre la préposition *avec* dans la mesure où les deux notions de saillance sont présentes dans la situation : saillance perceptive (gustative) et, lorsque ces « titres » sont affichés devant les restaurants, saillance commerciale.

Il faut toutefois noter que, dans les recettes de cuisine, la méronymie concernée n'est pas la même que celle qui est présente dans les petites annonces immobilières, les catalogues de jouets ou les sites touristiques. En effet, si l'on reprend la catégorisation de Winston et Chaffin (1987) en 6 classes :

- composant/objet,
- membre/collection,
- portion/masse,
- constituant/objet,
- activité/phase,
- zone/lieu,

on voit que la méronymie que nous avons identifiée dans les corpus étudiés concernait uniquement des cas composant/objet (cheminée/salle à manger, voiture/quatre roues motrices, église/portail...). La méronymie dont il est question dans les recettes de cuisine concerne, elle, la classe constituant/objet : une fois intégré au tout, le constituant ne peut plus être distingué (champignons/omelette, potiron/flan).

Cette piste mériterait d'être explorée pour expliquer la différence de fonctionnement entre *avec* et *à* mais il restera des cas où aucun argument ne peut expliquer le choix de l'une ou l'autre des prépositions (par exemple, on trouve des *à* dans des catalogues de vêtements : *robe à manches ballons*).

On le voit donc, le genre textuel peut avoir un rôle descriptif et prédictif pour expliquer le fonctionnement méronymique d'une préposition comme *avec*. Mais la dimension prédictive est à entendre comme caractérisant un fonctionnement langagier en lien avec une situation clairement identifiée. L'élargissement à des situations qui ne sont pas seulement identiques mais dont on suppose qu'elles ont des caractéristiques que l'on peut considérer comme similaires *a priori* reste, pour l'instant, plus délicate.

4.2. Le cas de la reprise anaphorique infidèle

La reprise anaphorique infidèle, examinée dans ses possibilités de « marquer » une relation, fonctionne de manière très différente de la préposition *avec* et le genre textuel n'y est pas impliqué de la même façon.

La reprise anaphorique infidèle est un procédé de progression discursive qui permet d'utiliser un nom différent du nom antécédent mais qui lui est co-référent (Apothéloz 1995 ; Corblin 1990 ; Kleiber 1990 ; Marandin 1986). Il est connu qu'entre ces deux noms (l'anaphorique et l'antécédent) une relation d'hyponymie peut s'instaurer comme dans :

(9) *Un chat est entré. Cet animal au pelage clair paraissait en bonne santé.*

Nous avons mis sur pied une expérimentation pour évaluer le rôle du genre textuel dans la présence d'une relation hyperonymique entre anaphorique et antécédent (Condamines, 2005b). L'hypothèse à l'origine de cette étude était que cette relation hyperonymique serait plus fréquente dans certains genres textuels, les manuels techniques ou scientifiques, que dans d'autres genres. En effet, on prétend souvent que, dans des domaines spécialisés, les relations conceptuelles sont plus stables que dans d'autres domaines. Nous pensions que cette relation discursive pourrait se mettre en place en s'appuyant sur la stabilité supposée de ces relations.

Un corpus a été construit contenant cinq textes :

- un manuel de géomorphologie (le même que dans l'étude précédente), GEO,
- un manuel technique : guide de planification (fourni par EDF), GDP,
- un autre manuel technique : *Méthodes et Outils de Génie Logiciel pour l'Informatique Scientifique* (Mougis), également fourni par EDF, Mougis,
- un ensemble d'articles extraits du *Monde Diplomatique* (CR-Rom, 1989), LMD,
- un roman de Maupassant : *Bel Ami*.

Notons que nous n'avons examiné que les seules anaphores démonstratives pour différentes raisons qu'il est difficile de détailler ici. La principale tient au fait que l'autre principal déterminant possible, le défini, engendre beaucoup plus de cas de déictiques, que, par définition, nous avons exclus de l'étude puisque celle-ci ne recherchait que les seuls cas de relation anaphorique.

Voici les résultats quantitatifs qui ont été obtenus (pour des résultats plus détaillés, voir (Condamines,, 2005b) :

	GEO	GDP	Mouglis	LMD	Bel A.
Nb de mots	206 000	148 000	45 100	110 700	170 200
Anaphore hyperonymique	26 %	32%	60%	19%	15,5 %
Anaphore non-hyperonymique	74 %	68%	40%	81%	84,5 %

Tableau 2. Résultats quantitatifs dans l'étude de l'anaphore infidèle

Il apparaît que, hormis pour le manuel *Mouglis*, la relation hyperonymique est peu utilisée dans la mise en place d'une anaphore infidèle, même dans le cas des manuels. L'hypothèse à l'origine de cette étude ne semble ainsi pas vérifiée. Une analyse plus approfondie (qualitative) permet toutefois d'avoir une idée plus précise du fonctionnement de ce type d'anaphore et de mettre en évidence un fonctionnement propre aux manuels. Tout d'abord, il faut bien comprendre que l'anaphore infidèle n'est pas un marqueur au même titre que *avec*. En effet, la relation entre anaphorique et antécédent n'est pas exprimée par le discours mais supposée connue du locuteur et de l'interlocuteur : elle est en fait présupposée. Dans l'exemple (9), la relation d'hyperonymie entre *animal* et *chat* n'est pas exprimée ; au contraire, le processus d'anaphorisation se construit sur son existence implicite. De ce fait, dans les corpus spécialisés, il sera parfois difficile pour un néophyte d'identifier la nature de la relation entre l'anaphorique et l'antécédent. Ainsi dans l'exemple suivant, extrait d'une revue spécialisée sur le vol à voile :

- (10) *Un écopage dissymétrique se rattrape facilement en recentrage/commande et n'interrompt pas le décollage. L'aile reste très légère aux commandes et aux élévateurs durant toute cette phase du début de vol.*

Il est difficile de décider si, entre *phase du début de vol* et *décollage*, il y a une relation d'hyperonymie (il y aurait donc d'autres phases en début de vol) ou de synonymie (phase de début de vol et décollage seraient équivalents).

Une autre difficulté tient à ce que, dans un certain nombre de cas, le nom anaphorique n'entretient pas une relation avec un autre nom antécédent mais avec toute une partie du texte précédent :

- (11) *La pression atmosphérique avait été évaluée il y a une trentaine d'années à 1.12 de celle de l'atmosphère terrestre. On a réduit cette appréciation car Mariner 4 a trouvé qu'elle équivalait à 6 milibars [...] (GEO).*

A la suite de Lerat (1981), nous avons appelé ces noms des « supplétifs ». Ils sont assez nombreux dans les différents textes étudiés :

GEO : 50 %,
 GDP : 55 %,
 Mouglis : 31,5 %,
 LMD : 64,5 %
 Bel Ami : 47 %.

Or, ces supplétifs ont un fonctionnement particulier dans les manuels : ils sont plus souvent des « classifieurs » que dans les autres textes.

Dans le cas où la relation est de nature hyperonymique, le nom-anaphorique est forcément un « classifieur », c'est-à-dire un nom qui peut constituer le sommet d'une hiérarchie taxinomique comme *pays, région, organisation* dans *Le Monde Diplomatique*. En revanche les supplétifs peuvent être soit des classifieurs, soit des non-classifieurs (nom de qualité, de caractérisation...) comme *accusation, démarche, exercice* toujours dans *Le Monde Diplomatique*. Il semble bien que, dans le cas des manuels, les supplétifs soient très souvent des classifieurs. Trois arguments semblent plaider en faveur de cette hypothèse :

- Certains noms en position d'anaphore, qui, hors contexte seraient identifiés comme des non-classifieurs sont utilisés, dans le texte, comme des classifieurs. C'est le cas de *responsabilité* dans Mougliis :

(12) *Quatre types de responsabilités sont associés au Plan de Gestion de Configuration : La rédaction et l'évolution du PGC .*

Cette responsabilité incombe au Responsable Assurance Qualité de niveau 1 en collaboration avec le Chef de projet. (Mougliis).

Cet extrait permet d'interpréter *rédaction du PGC* et *évolution du PGC* comme des hyponymes de *responsabilité* et donc de donner un rôle de classifieur à *responsabilité* (et pas seulement de caractériser).

- Dans les manuels, plusieurs noms sont répertoriés à la fois dans la rubrique « hyperonymes » et dans la rubrique « supplétifs ». Ainsi, dans Mougliis : *activité, aspect, phase, responsabilité, thème, transfert*, soit 6 noms sur 39 sont hyperonymes dans certains exemples et supplétifs dans d'autres. Cette situation montre que certains noms ont un statut très particulier dans ces manuels. Voici par exemple deux extraits, l'un dans lequel *activité* est hyperonyme, l'autre dans lequel il est supplétif.

(13) *Test du logiciel*

Cette activité consiste à exécuter les procédures de tests spécifiées pour qualifier le logiciel.

(14) *Les unités de configuration à modifier doivent être identifiées afin de pouvoir séquentialiser les modifications à apporter à un même composant. Le résultat de cette activité est la constitution du dossier de modification regroupant de façon ordonnée les demandes de modification affectées aux différents développeurs.*

- Enfin, la grande majorité des noms qui fonctionnent soit comme des hyperonymes soit comme des supplétifs dans les manuels techniques, sont les têtes de syntagmes nominaux dans ces mêmes textes. Prenons le cas de Mougliis. Sur 39 N (hyperonymes ou supplétifs), 35 apparaissent comme têtes de groupes nominaux dans le corpus, par exemple :

Action (de contrôle qualité, de gestion de configuration, de regroupement de fichiers...)

Espace (de stockage, de configuration, de référence logiciel, de développement, de test, de réception, de travail, de livraison, de gestion).

Or, on sait que si les groupes nominaux sont si nombreux dans les corpus spécialisés, c'est parce que la tête qui les constituent permet d'identifier une classe de noms qui, par l'ajout de modificateurs, se déclinent en autant d'hyponymes.

Ainsi, si l'hypothèse de l'anaphore infidèle comme marqueur d'hyperonymie est invalidée par cette étude, une autre, plus sûre, apparaît : l'anaphore infidèle pourrait être un marqueur d'hyperonyme. Plus précisément, les noms utilisés comme anaphoriques seraient plus souvent des classificateurs dans les manuels que dans des textes relevant d'autres genres. En d'autres termes, l'anaphore nominale infidèle pourrait être un marqueur d'hyperonyme et pas d'hyperonymie. Il faudrait alors utiliser des marqueurs d'hyperonymie plus classique comme [det N1 être dét N2 + modifieur] pour identifier les hyponymes. On aurait, pour rechercher les contextes correspondant à un tel marqueur, une entrée : le N hyperonyme, c'est-à-dire N2 dans le schéma et on chercherait un ou des N1 correspondant dans le corpus étudié.

Une fois encore, le fil conducteur qui a permis cette étude de l'anaphore infidèle : la recherche de marqueurs d'hyperonymie et l'étude du genre textuel a permis de préciser le fonctionnement de ce mécanisme discursif, par ailleurs largement étudié. Cette étude montre aussi que le rôle du genre textuel n'est jamais identifiable *a priori* et qu'il nécessite de mettre en place des analyses parfois longues et coûteuses. Enfin, même lorsque le genre textuel est pertinent, son identification n'est pas toujours aisée. En effet, selon l'objectif de l'étude et/ou le fonctionnement langagier que l'on souhaite étudier, la catégorisation en genres pertinents peut varier. Prenons l'exemple des petites annonces qui a été évoqué dans cet article. Seules les petites annonces immobilières sont pertinentes pour décrire le fonctionnement de *avec* ; en effet, les petites annonces de vente de voiture par exemple n'utilisent *avec* que de manière exceptionnelle. Les composants des automobiles à vendre sont, dans la plupart des cas, listés sans aucun marqueur. Mais si on s'attache à d'autres fonctionnements langagiers : longueur du message, présence d'ellipses, déficit de verbes... il est clair que les petites annonces en général, quel que soit l'objet à vendre ou à acheter, constituent un genre textuel pertinent.

5. Conclusion

Depuis 15 à 20 ans que la linguistique de corpus a émergé, mettant à disposition des textes en grande quantité, la réflexion en sémantique a été traversée de querelles scientifiques sur les approches possibles. Les uns, fervents défenseurs d'une approche introspective, prenaient les attestations disponibles sur support informatique comme autant de possibilités de confirmer (ou pas) leurs intuitions. Les autres, un peu sidérés par les quantités de données disponibles s'en sont surtout tenus à des approches quantitatives accompagnés (et c'est un élément déterminant) d'une réflexion sur la constitution des corpus.

La situation est en train d'évoluer et les approches pourraient se reconnaître complémentaires mais cette progression suppose que soient acceptés différents éléments :

- L'analyse sémantique ou même l'attribution d'un sens relève toujours d'une construction, c'est-à-dire d'une interprétation. Cela signifie qu'il y a une dimension de variation qui peut être importante en analyse sémantique. Le rôle du linguiste est alors de décrire, d'expliquer voire de prédire cette variation.

- Dès que les études sont faites à partir d'attestations réelles, elles doivent prendre en compte le fait que la situation dans laquelle ces attestations (des discours en fait) sont produites ne peut pas d'emblée être neutralisée. C'est un résultat de l'étude que de voir comment le contexte extra-linguistique peut être ou non, négligé dans la description.

- L'introspection est toujours présente, quelle que soit l'approche mise en œuvre ; qu'il s'agisse de catégoriser des situations ou des mots, l'intuition linguistique est toujours mobilisée, au tout début des études, quitte à ce que cette catégorisation soit ensuite remise en question.

- Le recours à la compétence de locuteur est elle aussi présente quel que soit le mode d'approche choisi : même dans le cas où les données étudiées ne sont pas produites par le linguiste, il est évident qu'il met en œuvre une compétence de locuteur qui intervient dans l'interprétation des données.

Le fait que la demande d'analyse de textes, surtout de la part d'organismes producteurs, soit très pressante ne doit pas effrayer les sémanticiens. En effet, cette demande peut permettre de poser des questions nouvelles qui ouvrent des perspectives inédites. C'est tout du moins le cas lorsque les analyses menées s'inscrivent dans une perspective de généralisation des résultats. La différence traditionnelle entre sémantique théorique et sémantique appliquée perd alors beaucoup de sa pertinence pour laisser la place à une sémantique située. Cette façon de voir peut permettre de comprendre comment se met en place l'interprétation sémantique et d'évaluer les possibilités de généraliser les modes d'interprétations.

<anne.condamines@univ-tlse2.fr>

Références

- Apothéloz D. (1995). Nominalisations, référents clandestins et anaphores atypiques. *Tranel* 23, 143-173.
- Aussenac-Gilles N. & Condamines A. (2004). Outils et méthodes pour l'amélioration de la visibilité des PME sur le web. Rapport de fin de contrat avec la Région Midi-Pyrénées, rapport interne IRIT n° 2004-28-R <http://www.irit.fr/ASSTICCOT/>
- Bouquet S. (2004). Linguistique générale et linguistique des genres. In S. Bouquet (ed.), *Les genres de la parole, Langages* 153, 3-14.
- Cadiot P. (1997). Avec, ou le déploiement de l'éventail. In C. Guimier (ed.), *Co-texte et calcul du sens*; Caen : Presses Universitaires de Caen, 135-155.
- Choi-Jonin I. (1995). La préposition *avec* : opérateur de (dé)composition. *Scolia* 5. 109-129.
- Condamines A. (2000). *Chez* dans un corpus de sciences naturelles : un marqueur de méronymie ? *Cahiers de Lexicologie*, 77, 165-187.
- Condamines A. (2005). Sémantique et corpus : quelles rencontres possibles ? In A. Condamines (ed.), *Sémantique et Corpus*, London, Hermes, 17-38.
- Condamines A. (2005). Anaphore nominale infidèle et hyperonymie : le rôle du genre textuel. *Revue de Sémantique et Pragmatique* 18.

- Condamines A. (2006). *Avec* et l'expression de la méronymie : l'importance du genre textuel. In G. Kleiber, C. Schnedecker & A. Thyssen (eds.), *La relation 'Partie-Tout'*. Leuven, Peeters.
- Corblin F. (1990). Typologie des reprises linguistiques : l'anaphore nominale. In Charolles M., Fisher S. & Jayez J. (eds.), *Le discours*, Nancy, Presses Universitaires de Nancy, 227-242.
- Cruse D.A. (1986). *Lexical Semantics*. Cambridge, CUP.
- Genette G. (1986). *Théorie des genres*. Paris, Le Seuil.
- Habert, B., Nazarenko, A. & Salem, A. (1997). *Les linguistiques de corpus*. Paris, A. Colin.
- Kleiber G. (1990). Sur l'anaphore démonstrative. In Charolles, M., Fisher, S. & Jayez, J. (eds.), *Le discours*, Nancy, Presses Universitaires de Nancy. 243-262.
- Lerat P. (1981). Les noms de relation. *Cahiers de Lexicologie*, 39-2, 55-65.
- Marandin J.M. (1986). *Ce* est un autre. L'interprétation anaphorique du syntagme démonstratif. *Langages* 81. 75-90.
- Lyons J. (1978). *Eléments de sémantique*. Paris, Larousse.
- Rastier F. (1987). *Sémantique interprétative*. Paris, PUF.
- Salanskis J.-M., Rastier F. & Scheps R. (1991). *Herméneutique : textes, sciences*. Paris, PUF.
- Slodzian M. (2000). L'émergence d'une terminologie textuelle et le retour du sens. In H.Béjoint & P.Thoiron (eds.), *Le sens en terminologie*, Lyon, PUL, 61-85.
- Swales J.M. (1990). *Genre Analysis, English in Academic and research settings*, Cambridge, CUP.
- Winston M.E., Chaffin R. & Hermann D.J. (1987). A taxonomy of part-whole relations. *Cognitive Science*, 11. 417-444.